

Gut mycobiota, serum anti-*Saccharomyces cerevisiae* antibody, and calprotectin concentrations in inflammatory bowel disease patients during infliximab therapy.

Rebecka Ventin-Holmberg

Master's thesis

University of Helsinki

April 2019

Tiedekunta — Fakultet/ — Faculty Faculty of Agriculture and Forestry <sup>a</sup> and Faculty of Biological and Environmental Sciences ( <sup>a</sup> coordination)		Maisterinohjelma – Magisters program Masters´ Programme Masters´ Programme of Microbiology and Microbial Biotechnology	
Tekijä — Författare — Author Rebecka Ventin-Holmberg			
Työn nimi — Arbetets titel — Title Gut mycobiota, serum anti- <i>Saccharomyces cerevisiae</i> antibody, and calprotectin concentrations in inflammatory bowel disease patients during infliximab therapy.			
Työn laji — Arbetets art — Level Master´s Thesis		Aika — Datum — Month and year April 2019	
Tiivistelmä — Referat — Abstract <p>Inflammatory bowel disease (IBD) is a globally increasing chronic disease, for which the pathogenesis still is unclear. The most common subtypes of IBD are Crohn's disease (CD) and ulcerative colitis (UC). It is widely known that, in addition to the genetics, an altered immune response against the gut microbiome plays an important role in the development of the disease. For the IBD patients, to whom conventional medication is not sufficient, the TNF-<math>\alpha</math> blocker infliximab, is given. However, about one third of the patients receiving infliximab treatment, do not respond to the drug, or lose response over time. Since there to this day are no reliable diagnostic markers available, the finding of such is of great importance. The goal of this study was to investigate possible markers for drug response in the gut mycobiota composition of IBD patients. The gut mycobiota composition of 72 IBD patients receiving infliximab was studied by MiSeq sequencing of fungal DNA from fecal samples, collected during one year. The sequencing data was analyzed using the <i>mare</i> package in R. In addition, anti-<i>Saccharomyces cerevisiae</i> antibody (ASCA) concentrations were measured from baseline serum samples by ELISA. Finally, calprotectin concentrations were measured from baseline and twelve weeks post infliximab serum samples by ELISA to study whether serum samples could be used instead of fecal samples for measuring calprotectin values.</p> <p>Results show an increase of the <i>Candida</i> and <i>Spiromyces</i> genera in the gut mycobiota of non-responding patients at baseline. At all timepoints, the <i>Spiromyces</i> genus was observed at a higher abundance, compared to the group of patients responding well or partially to the medication. Interestingly, the increase of <i>Candida</i> was seen only in Crohn's disease patients, when looking at the composition at all timepoints. ASCA values did not differ between the response groups. The serum calprotectin values did not correlate with fecal calprotectin, and serum calprotectin can thus not be used as a marker of gut inflammation. In conclusion, the gut mycobiota can offer predictive markers for drug response prediction to infliximab in IBD patients, which can with further studies offer a clinical diagnostic tool for prediction of drug response.</p>			
Avainsanat — Nyckelord — Keywords Gut mycobiota, IBD, Mycobiome, Microbiome, ASCA, Calprotectin, Infliximab			
Säilytyspaikka — Förvaringsställe — Where deposited <a href="https://ethesis.helsinki.fi/en">https://ethesis.helsinki.fi/en</a>			
Muita tietoja — Övriga uppgifter — Further information Supervision: Eija Nissilä and Päivi Saavalainen and Tekes/ Business Finland			
		Funded by: Sigrid Jusélius Foundation	

Tiedekunta — Fakultet/ — Faculty Agrikultur-forstvetenskapliga Fakulteten <sup>a</sup> och Bio- och miljövetenskapliga fakulteten ( <sup>a</sup> koordinerande fakultet)		Maisterinohjelma – Magisters program Masters' Programme Magisterprogrammet i mikrobiologi och mikrobiell bioteknik	
Tekijä — Författare — Author Rebecka Ventin-Holmberg			
Työn nimi — Arbetets titel — Title Tarmmykobiota, serum anti- <i>Saccharomyces cerevisiae</i> antikroppar och calprotectin koncentrationer hos patienter med inflammatorisk tarmsjukdom som får infliximab medicinering.			
Työn laji — Arbetets art — Level Pro Gradu-avhandling		Aika — Datum — Month and year April 2019	
Tiivistelmä — Referat — Abstract <p>Inflammatorisk tarmsjukdom (IBD) är en kronisk inflammation i tarmarna som är allt mer förekommande i hela världen. De vanligaste typerna av IBD är Crohn's sjukdom och ulcerös colit. En tydlig orsak till sjukdomen har inte blivit hittad, men man vet att IBD delvis orsakas av en ovanlig immunreaktion mot tarmens mikrobiom hos en genetiskt belagd person. Åt patienter, för vilka den vanliga inflammationshämmande medicineringen inte är tillräcklig, ges den biologiska medicinen infliximab. Upp till en tredjedel av patienterna som får infliximab får ingen läkemedelsrespons mot medicinen, antingen på grund av brist på respons från början, eller avfall av respons efter medicinering. Det är mycket viktigt att forska i prediktiva markörer, eftersom det i nuläget inte finns diagnostiska metoder för undersökning av respons mot infliximab. Målet med denna studie var att undersöka möjliga markörer för läkemedelsrespons i tarmmykobiotan hos IBD patienter. Tarmmykobiotans komposition undersöktes hos 72 IBD patienter, som får denna medicinering, genom MiSeq sekvensering av svamp DNA i avföringsprov tagna under ett år. Koncentrationen av anti-<i>Saccharomyces cerevisiae</i> antikroppar (ASCA) har mätts från serumprov tagna före medicineringen började genom ELISA. Slutligen mättes även koncentrationen av calprotectin från serumprov tagna före medicineringen började samt tolv veckor efter medicinering för att undersöka om dessa kunde användas istället för calprotectin värden mätta från avföring.</p> <p>I studien observeras en ökning av släkten <i>Candida</i> och <i>Spiromyces</i> i tarmmykobiotan hos gruppen av patienter för vilka infliximab inte fungerade i prover tagna innan påbörjad medicinering. En ökning av släktet <i>Spiromyces</i> var observerad även i prover tagna under medicinering, medan ökningen av <i>Candida</i> bara var observerad hos patienter som lider av Crohn's sjukdom i prover tagna under medicineringen. Det observerades ingen korrelation mellan calprotectinvärden mätta från serum och avföring, vilket betyder att serum calprotectinvärden inte fungerar som inflammationsmarkör för tarmen. Slutligen observerades inga skildaktigheter i ASCA-värden i de olika grupperna. Sammanfattningsvis visar studien att tarmmykobiotan erbjuder möjliga prediktiva markörer för läkemedelsrespons mot infliximab hos IBD patienter som kan användas i diagnostiken.</p>			
Avainsanat — Nyckelord — Keywords Mykobiota, IBD, Mykobiom, Mikrobiom, ASCA, Calprotectin, Infliximab			
Säilytyspaikka — Förvaringsställe — Where deposited <a href="https://ethesis.helsinki.fi/en">https://ethesis.helsinki.fi/en</a>			
Muita tietoja — Övriga uppgifter — Further information Supervision: Eija Nissilä och Päivi Saavalainen      Funded by: Sigrid Jusélius Foundation och Tekes/ Business Finland			

## INTRODUCTION

Inflammatory bowel disease (IBD) is a collective name of idiopathic diseases causing chronic inflammation in the gastrointestinal tract, of which Crohn's disease (CD) and ulcerative colitis (UC) are the most common [1]. Unclassified IBD (IBDU) is diagnosed in patients for whom the clinical decision is uncertain. IBD has a prevalence exceeding 0.3% in western countries, and the frequency has been increasing in recently industrialized countries located in Africa, Asia, and South America since 1990 [2, 3]. No defined pathogenesis has been established for IBD, but it is believed to derive from several environmental factors and particularly an inappropriate immune response against the gut microbiome in a host, who is genetically prone to IBD [4, 5].

There is no absolute cure for IBD, but the inflammation can be controlled with medication that induce and maintain remission. The treatment can for example consist of 5-aminosalicylic acid (5-ASA), and in some cases even corticosteroids [6]. Approximately 10–15% of the IBD patient receiving this treatment need further treatment to maintain the remission [7]. TNF- $\alpha$  is a proinflammatory cytokine, which is found to be associated with both CD and UC [8]. Infliximab is a chimeric monoclonal antibody, which binds to TNF- $\alpha$  thereby blocking its activity. Infliximab is a biological drug given to IBD patients to successfully maintain remission [9]. However, 10–40% of the CD patients and approximately 50% of the UC patients receiving infliximab do not respond to the medication. Additionally, up to 40% of the responders lose response over time [10].

The human body serves as host for a variety of different microorganisms, and our intestinal tract alone is colonized with up to  $10^{14}$  microorganisms, of which the majority resides in the colon [11]. These microorganisms make up the human microbiome and include bacteria, archaea, fungi, and viruses. Next-generation sequencing technology has enabled the unravelling of complex symbiosis between the microbiome and the host. In the human microbiota, bacteria are by far the most abundant, with approximately  $10^{11}$  bacterial cells compared to  $10^5$  fungal cells per gram of feces [12]. Naturally, this has led research to a focus on the bacteria. However, the fungal cells are 100-fold larger in volume compared to bacterial cells, signifying that these cannot be compared only by cell counts [13]. This suggests the importance of studying the mycobiota, meaning the fungal communities included in the microbiome.

There are multiple antibodies against several structures found on bacteria and fungi. Of these, the anti- *Saccharomyces cerevisiae* antibody (ASCA) is one of the most studied and is used as a serological marker for CD. ASCA is an antibody against oligomannosides in the cell wall of bakers' yeast [14], *S. cerevisiae*, which also is observed to be a member of the human gut mycobiota.

Calprotectin is a protein dimer of the subunits S100A8 and S100A9 and is present on the surface of macrophages and monocytes. The protein is considered to be an inflammatory marker, as its production increases during inflammation. Fecal calprotectin is a marker for IBD, and it is therefore used for the determining the disease activity [15]. Serum calprotectin is considered to derive from circulating leukocytes, rather than from the intestinal tract. According to this, high serum calprotectin levels may correlate with a more systematic inflammation, and not the local inflammation that is observed in IBD. Currently, few studies have taken serum calprotectin into

consideration when studying IBD, and no significant association has been observed between the fecal calprotectin and the serum calprotectin concentrations [16].

In this study, the aim was to discover new predictive markers for the drug response against infliximab in IBD patients, as no reliable diagnostic markers are available to this day. This was achieved by first characterizing the gut mycobiota in prospective IBD patients. The gut mycobiota composition was determined by using an internal transcribed spacer 1 (ITS1) rRNA MiSeq sequencing analysis of fungal DNA from stool samples collected before start of infliximab and throughout one year of treatment. In the sequencing, the ITS1 and ITS2 regions were targeted, amplifying the conserved, non-coding ITS1 region of fungal DNA, which previously has been used for this purpose [17]. As additional data supporting the study, the concentration of ASCA IgG and IgA were measured from baseline serum samples, and calprotectin concentrations were measured from baseline and twelve weeks post infliximab serum samples in prospective IBD patients using ELISA. As primary antibody in the calprotectin ELISA assay, 3406 was used in one combination and 3407 in the other. 3406 binds to the S100A8 subunit of the calprotectin protein complex and 3407 binds to the S100A8/A9 complex. As secondary antibody 3407 was used for both combinations. Thus, the 3407-3407 combination target the S100A8/A9 protein complex and the 3406-3407 target the S100A8 and S100A8/A9 complex. The calprotectin concentrations determined from serum samples were compared with those measured from stool samples of prospective IBD patients in connection with samples collection, to determine whether the serum calprotectin could be used as a similar inflammatory marker as fecal calprotectin is used. Finally, the MiSeq sequencing data was analyzed together with the additional data to find predictive markers for the drug response against infliximab.

## **METHODS**

### **Study design and subjects**

The cohort consisted of patients aged 13–64 years at the Department of Gastroenterology, Helsinki University Hospital, suffering from ulcerative colitis or Crohn's disease, for whom infliximab therapy was being initiated. The IBD patients had previously been diagnosed unresponsive to regular treatment with corticosteroids and immunosuppressants and the infliximab medication was therefore given to maintain remission. The response diagnosis to infliximab used in this study was assigned at twelve weeks post infliximab therapy, when the first response diagnosis is determined. Participants for the cohort were recruited during 2018 and 2019. Stool and serum samples were collected before starting infliximab therapy and two, six, and twelve weeks, as well as one year after the start of medication.

### **DNA extraction**

The fecal samples were collected and transported to the research facilities within approximately 8 hours, where they were stored at -80 °C until analysis. The DNA was extracted from stool samples using the repeated bead beating (RBB) method, which has previously been recommended [18]. The RBB method was developed for DNA extraction from bovine feces [19] and has been modified to extraction from rumen [20, 21]. Additional optimization has later been done for extraction of DNA from fecal samples [18]. In this study, the extraction method was further validated by comparison of the RBB method with and without the addition of proteinase K, which is included in most extraction methods used elsewhere [17]. The DNA extraction protocol used was the method optimized for fecal DNA extraction [18]. Briefly, the

fecal samples were partially thawed, weighed, and PBS was added to each sample. The samples were vortexed and transferred to bead beating tubes containing beads. After the addition of lysis buffer (1M Tris-HCl pH 8), the samples underwent the RBB procedure twice, including bead beating accomplished with the FastPrep®-96 (MP) instrument and incubation at 95 °C for ten minutes. Proteinase K was added using two concentrations after the first RBB procedure and incubated at 70 °C for 10 min. Finally, the DNA was extracted using the KingFisher FLEX (Thermo scientific) instrument. The DNA concentration of the extracted samples was determined using the high sensitivity Quant-iT™ dsDNA assay kit (Molecular Probes). The fluorescence was measured with the FLUOstar Optima plate reader (BMG LABTECH).

The samples extracted with both the RBB method and the RBB + proteinase K were run with qPCR to determine the difference between the methods. In qPCR, 1 ng of each primer was used [22], 10 µl of 2x BioRad iQ™ SYBR® Green Supermix, 5 ng DNA and addition of water to get the final volume of 25 µl for each reaction. The qPCR was performed in the BioRad CFX96 Touch™ instrument. No difference was observed in the DNA yield from the extraction, nor from the qPCR results and the study was therefore completed using the RBB extraction method without the addition of proteinase K.

### **Library preparation**

The fungal DNA was amplified with the PCR primer pair ITS1F (FWD, CTTGGTCATTTAGAGGAAGTAA) and ITS2 (REV, GCTGCGTTCTTCATCGATGC), which have been verified to target the conserved ITS region of fungal DNA [22] and proved to work more successfully than the other possible primer pairs [17]. The library preparation protocol was optimized by using different conditions in the PCR protocol for optimal product gain. In the final protocol, the library was prepared in a three-step PCR, where the ITS1 region first was amplified with the ITS1F and ITS2 primers mentioned. In the second PCR, the PCR product from the initial PCR reaction was amplified using ITS1F (FWD, ACACTCTTCCCTACACGACGCTCTTCCGATCTCTTGGTCATTTAGAGGAAGTAA) and ITS2 (REV, AGACGTGTGCTCTTCCGATCTGCTGCGTTCTTCATCGATGC) primers with Illumina adapters for MiSeq sequencing. The PCRs were carried out using a BioRad T100™ Thermal Cycler using the following conditions: denaturation at 98 °C for 60 s, 44 cycles at 98 °C for 10 s, 58 °C for 40 s, and 72 °C for 40 s, followed by a final extension time of 10 min at 72 °C. In the reaction mixture, 2x Phusion High-Fidelity PCR Master Mix with HF Buffer (ThermoFisher Scientific), 2 µl of each primer (5 µM), and 0.6 µl DMSO was used. Water was added to get a final volume of 20 µl for each reaction. For the initial PCR reaction, 5 ng DNA was used. In the following PCR reaction, 4.4 µl of the initial PCR product was used directly as template. After the second PCR, the product was purified using AMPure XP beads (Beckman Coulter, Copenhagen, Denmark) according to the 16S metagenomic library prep guide using 0.8x concentration of beads [23]. The purified products were analyzed using LabChip® GX Touch™ nucleic acid analyzer. The DNA concentrations of the samples were measured with the Quant-It PicoGreen dsDNA Assay Kit. The absorbance was measured using the FLUOstar plate reader. In the third PCR reaction, barcodes for sequencing were added to the samples by index PCR, primers previously published [24]. The PCR products were purified with AMPure XP beads, as previously, and analyzed with LabChip. The concentrations were measured again as previously. Finally, the samples were pooled for MiSeq sequencing by adding 10 nM of each sample to the pool.

## Analysis of sequencing data

The MiSeq sequencing data was preprocessed using the *mare* package (Microbiota analysis in R made easily, [25]) in R [26]. Only forward reads were included in the analysis due to poor quality of reverse reads. The processing of the forward reads was done with the “ProcessReads” function included in the *mare* package and the default parameters in *mare* were used. The reads were truncated to 150 nt to exclude the low quality region of the reads. Reads below the abundance of 0.00001% were discarded. USEARCH 8 [27] was used to dereplicate reads and to remove chimeras for annotation using the *unite* database for fungi [28]. Low quality reads were discarded from further analysis. The reads were clustered to operational taxonomic units (OTUs) using USEARCH8 [27], after which richness and diversity [29] were assigned for the samples. Further analysis was completed using visualization tools included in *mare* [25].

## ASCA IgG/IgA ELISA assay

IgG and IgA anti-*Saccharomyces cerevisiae* antibody (ASCA) concentrations were measured from IBD serum samples using a commercialized ELISA assay (ASCA IgG/IgA ELISA, Demeditec Diagnostics GmbH). The serum samples were diluted to 1:100, as recommended by the manufacturer, and were analyzed as singlicates for both IgG and IgA ELISA assays. Plates were read at 405 nm by the Hidex Sense Microplate Reader. The OD 405 nm values were analyzed by comparison to the standards included in the kit. Values exceeding or equal to 10 U ml<sup>-1</sup> were considered as positive outcomes, as recommended by the manufacturer.

## Calprotectin ELISA assay

The calprotectin concentration of IBD patient serum samples was measured with the ELISA sandwich method using two combinations of antibodies (Anti-human Calprotectin 3406 and 3407 SPTN-5, Medix Biochemica) to capture the calprotectin. The two antibody-combinations used were: 3407-3407 and 3406-3407, as recommended by the manufacturer. The samples measured in the ELISA assay were IBD serum samples taken before the start of infliximab therapy and twelve weeks after infliximab therapy and the samples were analyzed as duplicates. As standard, Recombinant Calprotectin Antigen (Medix Biochemica) was used starting from a concentration of 300 ng ml<sup>-1</sup> and diluted 1:3 in blocking buffer (TRIS-CaCl<sub>2</sub>-0.5% BSA, pH 8). One of the samples had previously been determined to have a high calprotectin concentration and was used as a control.

Shortly, the ELISA plates were coated with 100 µl primary antibody (2 µg ml<sup>-1</sup>, dilution in PBS) and incubated at +4 °C overnight. The fluid was then discarded, the wells were washed with 250 µl of washing buffer (100 mM TRIS-10 mM CaCl<sub>2</sub>-0.05% Tween-20, pH 8) once, blocked with 200 µl of blocking buffer (100 mM TRIS-10 mM CaCl<sub>2</sub>-0.5% BSA, pH 8) and incubated for approximately 1 h at room temperature. The plates were washed twice with 250 µl of washing buffer, the serum samples were diluted to 1:20 in blocking buffer and 100 µl of the dilution was added to each well. The plates were then incubated for approximately 1 h at room temperature with rotation. The plates were washed three times with 250 µl washing buffer, the secondary antibody was diluted to 1:3000 in blocking buffer and 100 µl was added to each well. The plates were then incubated as after the addition of serum. The plates were washed five times with 250 µl washing buffer and 100 µl substrate solution (para-nitrophenylphosphatase dissolved in 1M diethanolamine, 0.5 mM MgCl<sub>2</sub>) was added. When the colors changed to a linear range, the reaction was stopped using 100 µl 1M NaOH, and the

plates were finally read at 405 nm with Hidex Sense Microplate Reader. The OD 405 nm values were compared to those of the standards. The results were analyzed by first normalizing the variation between the plates by calculating the average of the control of each plate separately and adjusting these to the same level. The sample values were then adjusted by this factor.

## RESULTS

### Patient characteristics

A total of 73 IBD patients were included in this study, of whom 25 were diagnosed with CD, 46 with UC, and 2 with IBDU. Of the total number of patients, 53% responded well to infliximab treatment, 26% responded partially, and 14% were not responsive at twelve weeks after the start of infliximab therapy. Patients in the non-responding group that underwent colectomy surgery, due to no sustained remission, were included. The number of patients included varied in the different assays (Table 1). Additional patient characteristics are presented in Table 1.

**Table 1.** Basic characteristics of IBD patients

Characteristics	n (%) or Mean (Min-Max)
No. of patients	73
Male	43 (58.9)
No. of patients for	
Mycobiota characterization	72
ASCA IgG/IgA ELISA	69
Calprotectin ELISA	71
Age	
At diagnosis of IBD	28 (11-52)
At the time of analysis	37 (13-64)
Diagnose	
Crohn's disease	25 (34.2)
Ulcerative colitis	46 (63.0)
Unclassified IBD	2 (2.7)
Drug response (week 12 post IFX)	
Responders	39 (53.4)
Partial responders	19 (26.0)
Non-responders*	10 (13.7)
Disease duration (years)	9 (1-41)

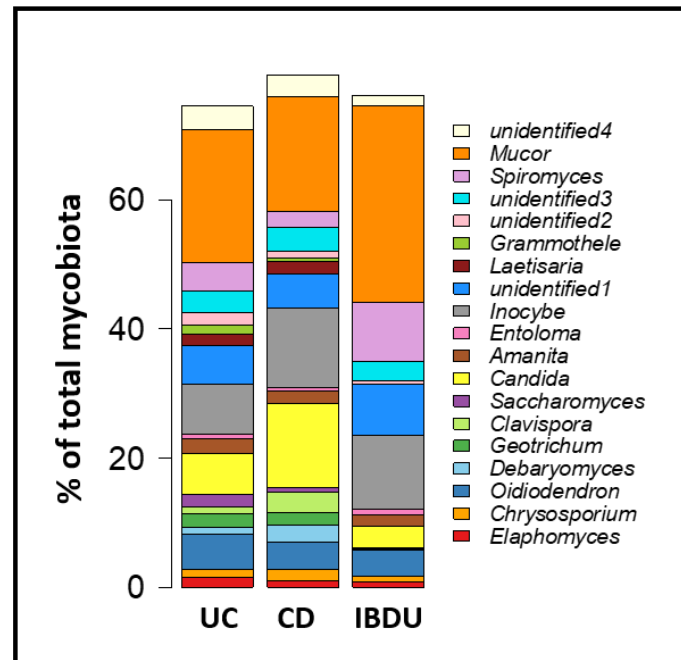
\*Patients that underwent colectomy before evaluating the drug response were included in the non-responders.

### Mycobiota characterization results

In this study, the gut mycobiota in IBD patients was characterized through Illumina MiSeq sequencing to detect possible markers for drug response to infliximab. The total number of reads in each sample before preprocessing using the mare package [25] was at an average of approximately 54 700 reads. After processing, by using only forward reads, the average decreased to 4700 high quality reads in each sample. The two main phyla of fungi observed in

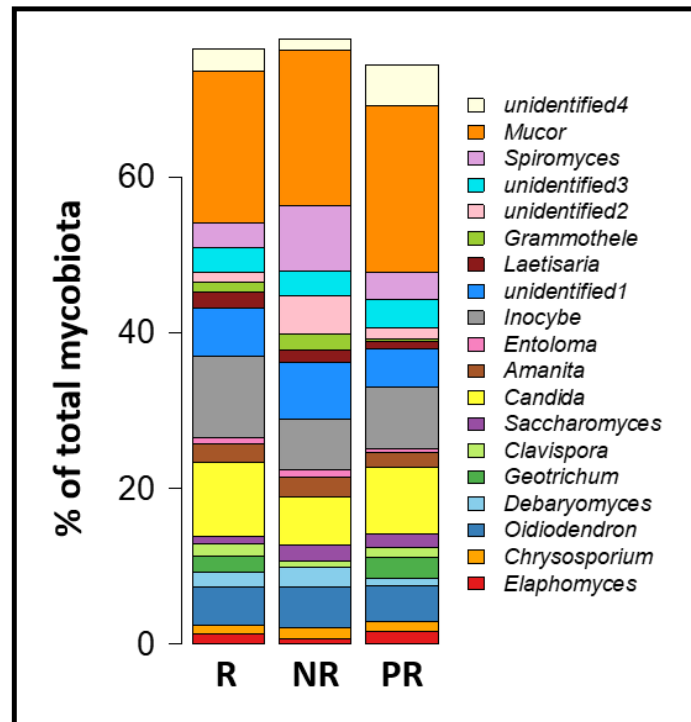


the mycobiota of IBD patients, when assigning taxonomy using the unite database [28], were Ascomycota and Basidiomycota. The most abundant genus in the gut mycobiota was *Mucor* for all IBD subtypes (Fig. 1). Other fungi present were of the genera *Spiromyces*, *Grammothele*, *Laetisaria*, *Inocybe*, *Entoloma*, *Amanita*, *Candida*, *Saccharomyces*, *Clavispora*, *Geotrichum*, *Debaryomyces*, *Oidiodendron*, *Chrysosporium* and *Elaphomyces* (Fig. 1). In Crohn's disease patients, *Candida* was more abundant, whereas *Saccharomyces* was less abundant, compared to the other IBD subtypes (Fig. 1).



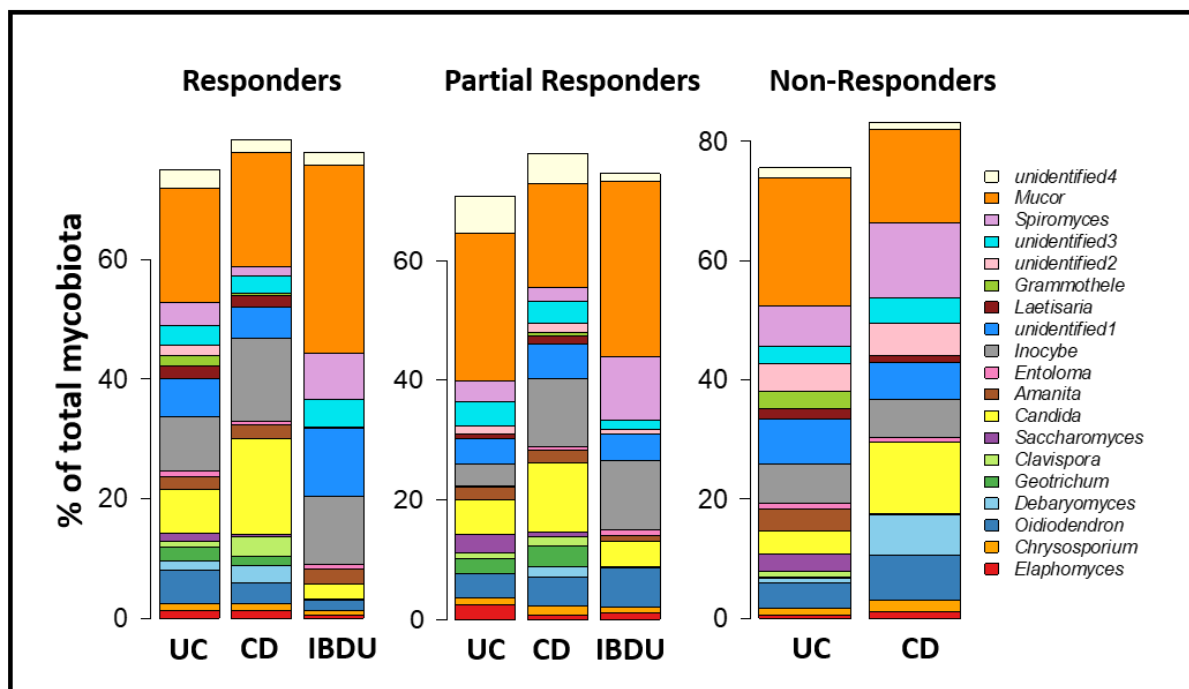
**Fig. 1.** Gut mycobiota composition in IBD subtypes UC (ulcerative colitis), CD (Crohn's disease) and IBDU (unclassified IBD). The plot presents the most abundant genera (mean abundance = 0.01), showing the relative abundance of the genera in each IBD subtype. The genera are color-coded and shown on the right side of the panel. The unidentified1-4 refer to four fungal taxa not assigned by the unite database.

The gut mycobiota was studied according to response against infliximab therapy to compare the compositions between the groups. The same genera of fungi can be observed to be present in both responders and partial responders, part from the *Clavispora* genus, that is not present in the group of partial responders (Fig. 2). In the group of non-responders, the genus *Geotrichum*, present in both responders and partial responders, is not observed and the genus *Spiromyces* is more abundant compared to the other groups (Fig. 2).



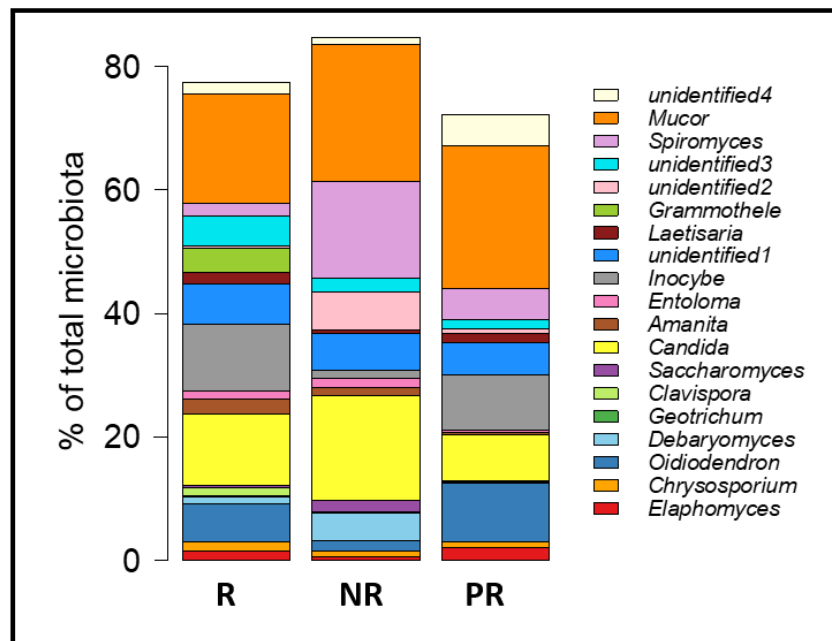
**Fig. 2.** Gut mycobacteria composition according to infliximab treatment response. R represents response, NR no response and PR partial response to infliximab therapy. The drug response was assigned to the patients at twelve weeks after the initiation of infliximab therapy. The plot presents the most abundant genera (mean abundance = 0.01), showing the relative abundance of the genera in the response groups. The genera are color-coded and shown on the right side of the panel. The unidentified1-4 refer to four fungal taxa not assigned by the unite database.

The gut mycobacteria at genus level did follow the same pattern in the variation between IBD subgroups, when divided further according to response against infliximab (assigned at twelve weeks post infliximab therapy), apart from the genus *Spiromyces*, which was more abundant in Crohn's disease patients not responding to infliximab, compared to the same IBD subtype that did respond, or responded partially to infliximab (Fig. 3).

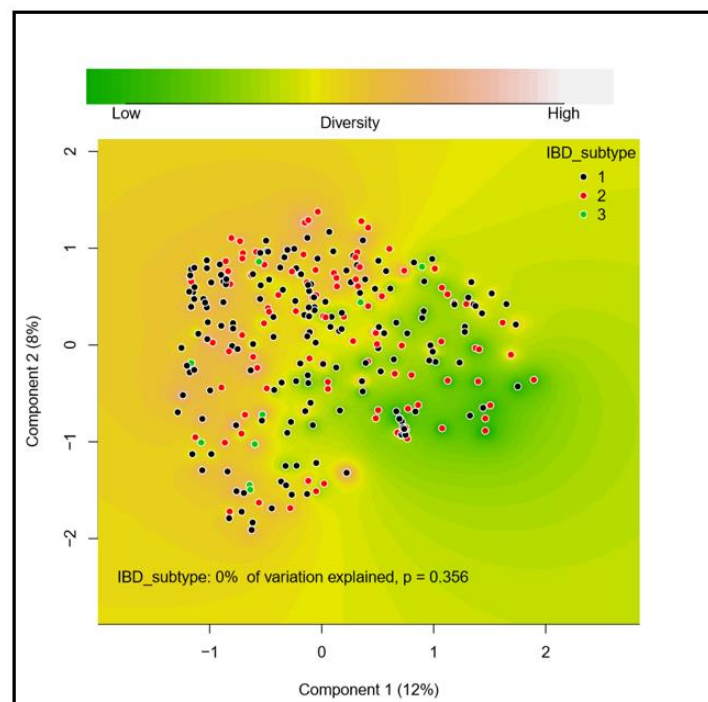


**Fig. 3.** Gut mycobacteria composition according to infliximab therapy response in IBD subtypes UC (ulcerative colitis), CD (Crohn's disease) and IBDU (unclassified IBD). The plot presents the most abundant genera (mean abundance = 0.01), showing the relative abundance of the genera for the response groups. The drug response was assigned to the patients at twelve weeks after the initiation of infliximab therapy. The genera are color-coded and shown on the right side of the panel. The unidentified1-4 refers to four fungal taxa not assigned by the unite database.

The possible markers in the gut mycobacteria were further investigated by studying the mycobacteria composition at the timepoint before start of infliximab therapy for the response groups. At baseline, the non-responders had a higher abundance of the genera *Candida* and *Spiromyces*, a lower abundance of the genus *Inocybe*, and no *Oidiodendron* was present, compared to the responders and partial responders (Fig. 4). No difference in mycobacteria composition according to IBD subtypes was observed when analyzing the subtypes using a principal coordinates analysis plot (Fig. 5).



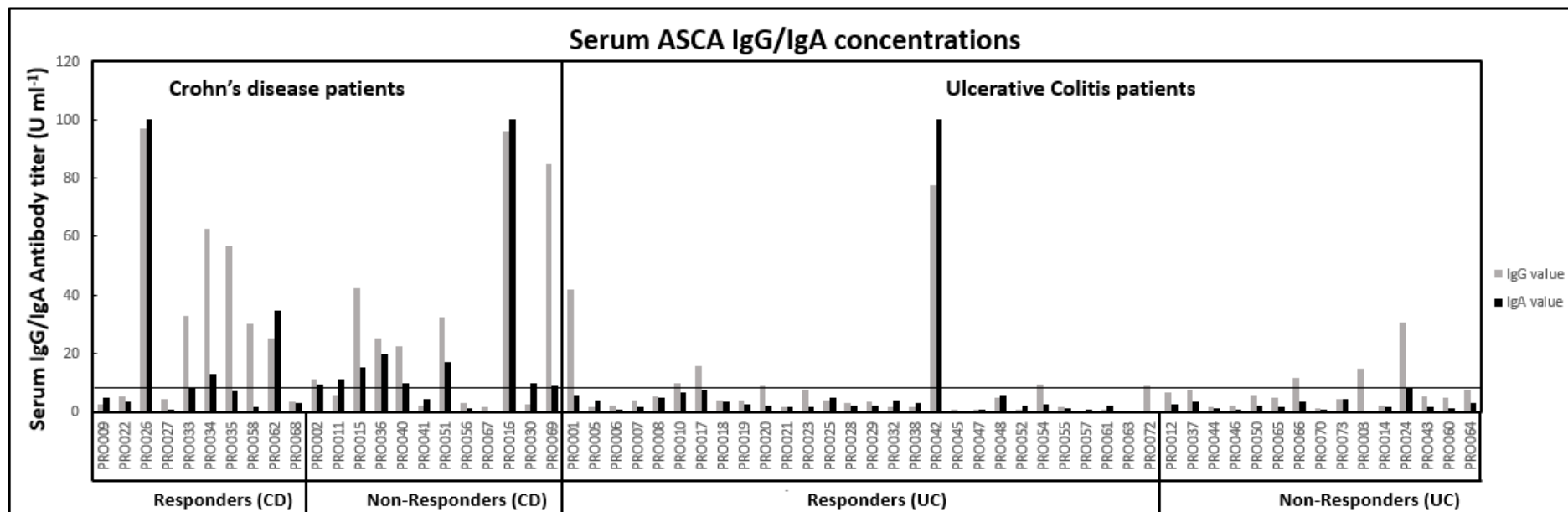
**Fig. 4.** Gut mycobiota composition of IBD patients at baseline according to infliximab treatment. R represents response, NR no response and PR partial response to infliximab therapy. The drug response was assigned to the patients at twelve weeks after the initiation of infliximab therapy. The plot presents the most abundant genera (mean abundance = 0.01), showing the relative abundance of the genera in the response groups. The genera are color-coded and shown on the right side of the panel. The unidentified1-4 refers to four fungal taxa not assigned by the unite database.



**Fig. 5.** Principal coordinates analysis (PCoA) plot according to IBD subtypes based on Bray-Curtis dissimilarity. The background color indicates interpolated values of diversity. The IBD subtypes are number-coded on the right side of the panel where 1 equals UC, 2 equals CD, and 3 equals IBUD.

### **ASCA IgG/IgA ELISA assay results**

The association between ASCA concentrations and response to infliximab treatment was studied to find differences between the response groups. This was achieved by analyzing IgG and IgA antibodies against ASCA in baseline serum samples from 69 prospective IBD patients using ELISA. Positive IgG values were observed in 29% and positive IgA values in 14% of the total number of patients. In patients suffering from Crohn's disease, the positive values were 59% of the IgG and 36% of the IgA values. In ulcerative colitis patients, however, the positive outcome was 14% of the IgG and 2% of the IgA values (Table 2).



**Fig. 7.** Serum ASCA IgG/IgA concentrations of IBD patients. In the plot, the ASCA values are presented for each IBD subtype (Crohn's disease (CD) and ulcerative colitis (UC)) and also according to response against infliximab. The drug response was assigned at twelve weeks post infliximab therapy. The unclassified IBD patients, and the patients with undefined drug response at week 12 post infliximab were excluded from the ASCA IgG/IgA data. Values exceeding 10 U ml<sup>-1</sup> were considered positive and this is indicated with a line in the figure.

**Table 2.** Serum ASCA IgG/IgA values of IBD patients at baseline

Characteristics	No. of patients	IgG positive	IgG negative	IgG pos (%)	IgA positive	IgA negative	IgA pos (%)
<b>All samples</b>	65	19	46	29.2	9	56	13.8
<b>Crohn's disease</b>	22	13	9	59.1	8	14	36.4
Response (CD)	10	6	4	60.0	3	7	30.0
Partial response (CD)	9	5	4	55.5	4	5	44.4
No response (CD)	3	2	1	66.7	1	2	33.3
<b>Ulcerative colitis</b>	43	6	37	14.0	1	42	2.0
Response (UC)	28	3	25	10.7	1	27	3.6
Partial response (UC)	9	1	8	11.1	0	9	0.0
No respons (UC)	6	2	4	33.3	0	6	0.0

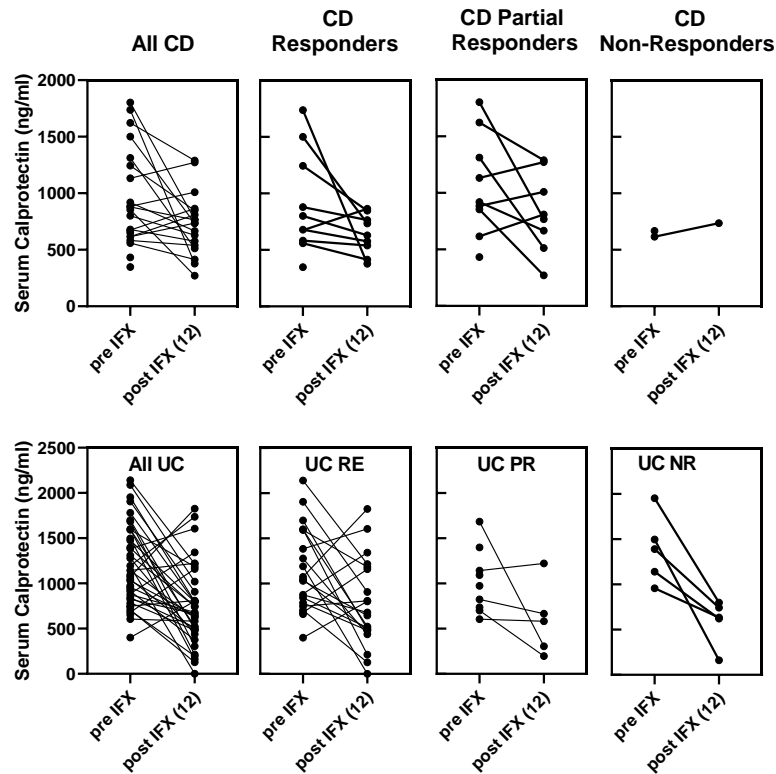
IFX response = response to infliximab therapy at twelve weeks post treatment. IgG pos (%) = the percentage of positive IgG values. IgA pos (%) = the percentage of positive IgA values.

For finding possible connections between concentration of ASCA and the presence of *S. cerevisiae* in the gut mycobiota, the association between ASCA values and the abundance of *Saccharomyces* family in the mycobiota was explored. However, no correlation between ASCA IgG/IgA values and Saccharomycetaceae or Saccharomycetales relative abundances (calculated from read counts of organized family table created by the R package mare [25]) were observed (Spearman test,  $r = -0.020$ ,  $p = 0.871$  for IgG, and  $r = -0.173$   $p = 0.265$  for IgA,  $n = 68$ ).

### Calprotectin ELISA assay results

In the calprotectin ELISA assay two antibody combinations were used for measuring the serum calprotectin by targeting the S100A8/A9 protein complex of calprotectin in baseline samples and 12-week post infliximab samples in 71 prospective IBD patients. This was done to investigate the correlation between serum calprotectin and fecal calprotectin values. For antibody combination 3406-3407, the serum calprotectin values were higher ( $P=0.0124$ ) before the start of infliximab therapy in CD patients ( $933 \pm 428$  ng ml<sup>-1</sup> (mean $\pm$ standard deviation)) than 12-weeks post infliximab ( $727 \pm 275$  ng ml<sup>-1</sup>). The values follow the same pattern for UC patients ( $P<0.0001$ ) with significantly higher values ( $1167 \pm 437$  ng ml<sup>-1</sup>) before treatment than post treatment ( $746 \pm 442$ ). In the only unresponsive CD patient with serum samples collected at both timepoints, the serum calprotectin value was higher 12-weeks post infliximab than before the start of medication for both antibody combinations used in ELISA (Fig. 8).

### Serum calprotectin ELISA antibody combination 3406-3407

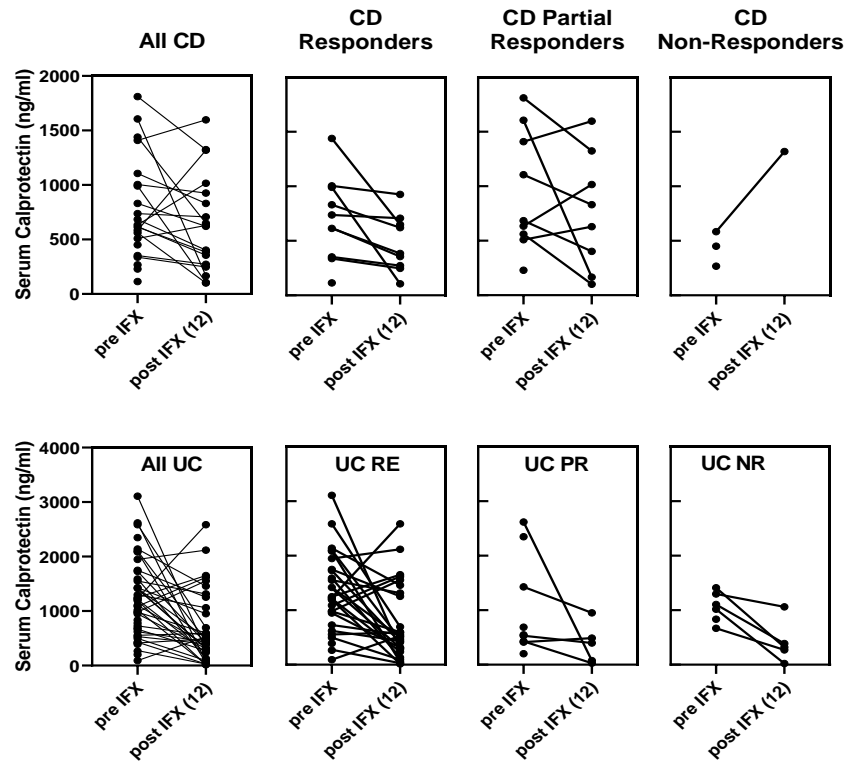


**Fig. 8.** Serum calprotectin values of IBD patients before and after infliximab therapy measured by ELISA using 3406-3407 antibody combination. All CD: all Crohn's disease patients. CD responders: all Crohn's disease patients responding, CD partial responders: all Crohn's disease patients responding partially, and CD non-responders: all Crohn's disease patients not responding to infliximab at twelve weeks post the start of treatment. Identically, All UC: all ulcerative colitis patients, UC RE: ulcerative colitis patients responding to treatment, UC PR: ulcerative colitis patients responding partially to treatment, UC NR: ulcerative colitis patients not responding to treatment. Pre IFX shows the calprotectin values before the start of infliximab therapy and post IFX (12) shows the calprotectin values after twelve weeks post infliximab treatment.

Similarly, for the antibody combination 3407-3407 the serum calprotectin values were higher in CD patients ( $p = 0.0597$ ) before start of infliximab therapy ( $768 \pm 464$ ) than after twelve weeks of infliximab treatment ( $648 \pm 447$ ). Again, the UC patients had significantly higher serum calprotectin values ( $p = 0.0002$ ) before start of infliximab therapy ( $1182 \pm 718$ ) than twelve weeks post therapy ( $662 \pm 641$ ). Identically as for the previous antibody combination, the only CD non-responder from whom serum samples has been collected at both timepoints showed higher post-infliximab value than before start of treatment (Fig. 9).



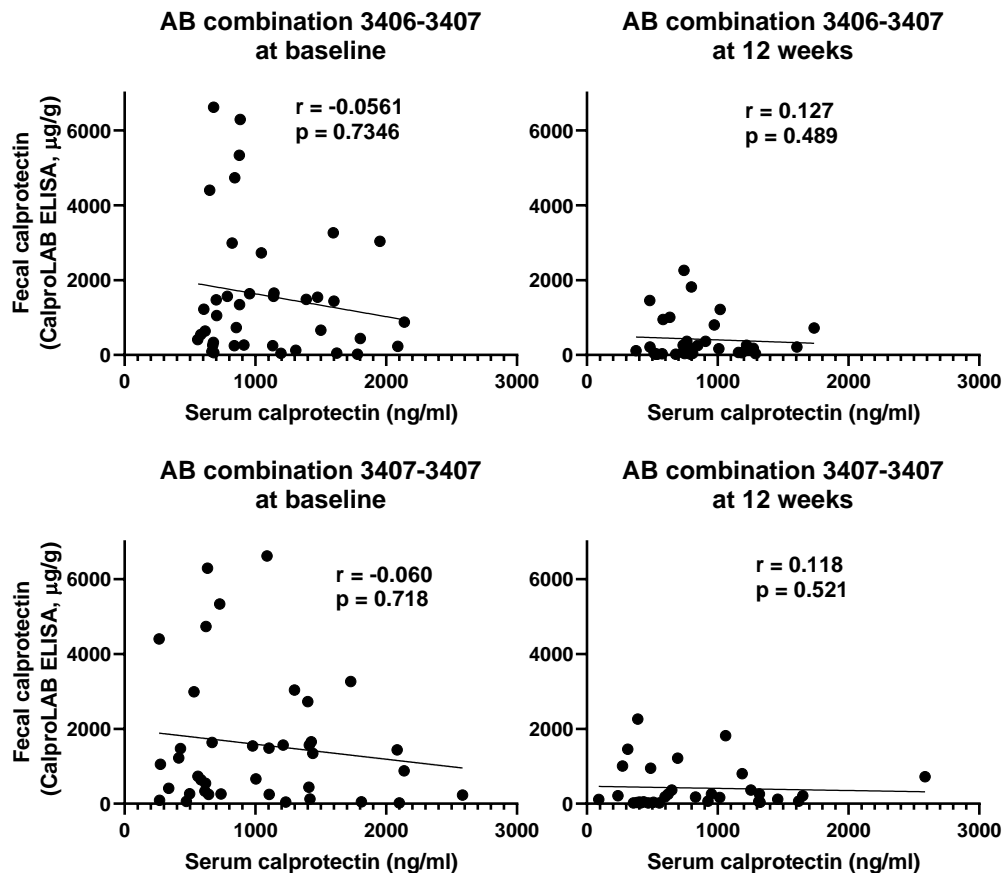
### Serum calprotectin ELISA antibody combination 3407-3407



**Fig. 9.** Serum calprotectin values of IBD patients before and after infliximab therapy measured by ELISA using 3407-3407 antibody combination. All CD: all Crohn's disease patients. CD responders: all Crohn's disease patients responding, CD partial responders: all Crohn's disease patients responding partially, and CD non-responders: all Crohn's disease patients not responding to infliximab at twelve weeks post the start of treatment. Identically, All UC: all ulcerative colitis patients, UC RE: ulcerative colitis patients responding to treatment, UC PR: ulcerative colitis patients responding partially to treatment, UC NR: ulcerative colitis patients not responding to treatment. Pre IFX shows the calprotectin values before the start of infliximab therapy and post IFX (12) shows the calprotectin values after twelve weeks post infliximab treatment.

No correlation was observed at baseline between serum calprotectin (antibody combinations 3406-3407) and fecal calprotectin measured with reference ELISA (CalproLAB), nor at 12 weeks after the start of infliximab therapy (Fig. 10.). The same pattern was determined for antibody combination 3407-3407 when comparing with fecal calprotectin measured with reference ELISA (CalproLAB) (Fig. 10.). Similarly, no correlation ( $p > 0.05$ ) was observed at the same timepoints between the serum calprotectin (ab combinations 3406-3407 and 3407-3407) and fecal calprotectin measured at HUSLAB (data not shown).

## Correlations between serum and fecal calprotectin



**Fig. 10.** Correlation between serum and fecal calprotectin values in IBD patients at baseline and 12-weeks post infliximab. In the plot, the correlations determined using Spearman correlation test are shown and the abbreviation AB represents antibody.

## DISCUSSION

IBD is a common disease in the industrialized countries, but in recent years its prevalence has increased, and it has become a global health care problem [2, 3]. Although, no exact pathogenesis has been determined for IBD, it is known that genetics, gut microbiome and environmental factors play a major role in the development of the disease [4, 5]. The biological drug infliximab is given to the patients for whom the regular treatment is ineffective. However, approximately one third of the patients suffering from Crohn's disease, and up to 50% of ulcerative colitis patients either do not respond to the treatment or lose response over time [10]. This is a concern, not only because of the high costs that the medication requires, but also because of the unnecessary side-effects that the ineffective drug causes to the patients. Finding predictive markers for drug response to infliximab is therefore of great importance. In the current study, this was achieved by characterizing the gut mycobiota of prospective IBD patients suffering from ulcerative colitis or Crohn's disease by MiSeq sequencing. The cohort population in this study had a relatively large patient number, which allows for detailed experimental analysis. As supportive data, ASCA IgG and IgA values were measured from baseline serum samples, and calprotectin values were measured using two antibody

combinations from baseline and twelve weeks post infliximab treatment serum samples with ELISA. Previously, predictive markers for drug response to infliximab have been found when studying the gut microbiota [30, 31], but to this day, no previous studies have investigated the possible role of the mycobiota in drug response against infliximab. The findings of this study revealed that there is a shift in the gut mycobiota composition in the group of patients that do not respond to the infliximab therapy compared to the group of patients responding well, or partially, to the drug.

The detected fungi of the gut mycobiota in IBD patients belonged to the phyla Ascomycota or Basidiomycota. Identically, it has been observed that IBD patients have a mycobiota composition consisting of these phylum [32, 33]. In this study, due to poor quality of reverse reads, only forward reads were included in the analysis and therefore the mycobiota was studied only at genus-level. As previously published; the most common genera in the gut mycobiota are *Saccharomyces* and *Candida* [17, 34, 32]. According to the results in this study, the most abundant genus was *Mucor*, the second most abundant genus was either *Inocybe* or *Candida*. Of the most abundant genera detected in this study, *Saccharomyces*, *Debaryomyces*, *Candida*, *Mucor*, and *Debaryomyces* have been observed in multiple previously published studies [35, 36, 32, 34, 17]. Additionally, the genera *Amanita*, *Clavispora*, and *Geotrichum* have been detected in less than two published studies as previously reviewed [35]. Finally, in the most abundant genera present in the mycobiota of IBD patients, four unidentified genera were not identified in the database, and therefore remained unidentified in the results.

In the gut mycobiota, an increased abundance of *Spiromyces* was observed at both baseline and when including all timepoints, especially in patients suffering from Crohn's disease. This indicates possible markers in the species of the *Spiromyces* genus. Before the start of infliximab treatment, the composition between the group of patients not responding to infliximab differed from the group responding well or partially to infliximab. The non-responding groups had higher abundance of *Candida* and *Spiromyces* and lower abundance of *Inocybe* and *Oididendron* in comparison to the other groups. The infliximab treatment is known to down-regulate inflammation in the gut and thereby restoring the normal gut epithelium [37], which suggests a possibility of altering the microbiome composition of the gut. The difference in the gut mycobiota composition between response-groups before the start of infliximab therapy observed in this study could indicate that the species of the gut mycobiota in non-responding individuals somehow prevents the drug from restoring the gut to normal, thereby making the drug inefficient.

The ASCA IgG and IgA values were measured from baseline serum samples in IBD patients by ELISA to study whether there was a difference between the group responding well to infliximab therapy and the group of non-responders. No difference between responders and non-responders was observed, but the patients suffering from Crohn's disease had higher ASCA values than the patients suffering from ulcerative colitis. These results are according to previously published data [14], and because of this fact, ASCA values are used as a diagnostic tool for Crohn's disease.

Fecal calprotectin concentrations are already used inflammatory marker in the gut. The serum calprotectin levels were explored to study whether serum calprotectin values can be used as a similar marker of gut inflammation as this would provide a more convenient sampling method for both the patient and diagnostics. Serum calprotectin levels were measured by ELISA for comparison with the fecal calprotectin levels already measured with reference ELISA. There

was no correlation observed between the fecal calprotectin concentrations and the serum calprotectin levels neither at baseline nor at twelve weeks post infliximab and this has also previously been observed, using different antibodies [16]. However, the calprotectin concentrations did decrease when comparing the levels at twelve weeks to the levels at baseline. These results follow the pattern of fecal calprotectin levels, suggesting the serum calprotectin levels working more as a marker of systematic inflammation, as previously reported [38, 39, 40], and, thus, cannot be used as an inflammatory marker of the gut.

As previously reviewed [36], the choice of primers, and particularly the choice of database and pipelines for analysis of data highly impact the results of a mycobiome study. The primers amplifying the ITS1 region was used in this study, and as previously stated [36, 17], this should be working sufficiently for this particular purpose. This primarily leaves the choice of analysis pipeline and database for further investigation. In this study, the *mare* package [25] and the UNITE database [28] was used for analysis of sequencing data. Different pipelines have been compared for the analysis of 16S sequencing data previously [41] and no difference was observed. In the *mare* package [25], comparison between *mare* and QIIME [42] is included, and *mare* has been proved to be work better for 16S data analysis specifically. The *mare* package was chosen based on this, in addition to the inbuilt statistical analysis tools, but it is developed for analysis of bacterial data, and therefore it might not be optimal for the analysis of ITS sequencing data. In addition, databases for assigning taxonomy should be compared for optimal results. This could be the reason to the four unidentified genera appearing in the plots, and to the most abundant genera in this study, that has not been detected in other studies of the human gut mycobiota. Another alarming outcome in the processing phase using the *mare* package was the massive decrease of reads passing through the filters. This is partially due to high levels of A-nucleotides added to the short reads in the sequencing due to differing lengths of reads, but it is my conclusion that the inbuilt filtering in *mare* is too strict for ITS-processing thereby causing the massive decrease of read numbers. Furthermore, the diverging results gained in this study suggests reason to question the trustworthiness of the output until comparison between pipelines and databases have been completed. After re-evaluation of the analysis methods, the mycobiota results should be analyzed together with the ASCA results to verify the correlation between Saccharomycetales and ASCA values.

Other pipelines that can be used for analysis of ITS sequencing data are the QIIME pipeline [42, 34], DADA2 pipeline [43], BROCC pipeline [44], and CloVR-ITS pipeline [45]. The *mare* package [25] is proved to work well for analysis of 16S data, as previously mentioned [46] and in the *mare* package, USEARCH8 [27] is used for assigning OTUS and processing reads. In other pipelines, such as DADA2 [35] and QIIME [42], this is achieved using different tools. Both QIIME and DADA2 pipelines offer a specific workflow for processing of ITS data, which seems logical because of the difference between 16S and ITS sequencing data. The results gained by using the UNITE database [28] for assigning taxonomy should also be compared by using other databases such as BLAST [47]. To conclude, I believe that ITS data must be processed differently compared to 16S data, and therefore, the usage of the *mare* package in such is not optimal, and an ITS specific pipeline is therefore crucial for the analysis.

In conclusion, differences in the gut mycobiota in the non-responding group, compared to the group responding well when studying the composition at all timepoints, and particularly at baseline, were observed. Based on this study, the variations of composition presented in the mycobiota could present predictive markers for drug response to infliximab therapy. More

detailed studies are required to specify the study, including using a pipeline optimized for the analysis of ITS data, studying the gut microbiota, virobiota and eucaryota from the same patients studied here, and further investigate the interaction between the microorganisms, of which the importance previously has been stated [48]. In addition, the gut mycobiota should be studied at species-level to further specify the markers and finally, to gain a more detailed composition of the fecal panel that would provide a predictive diagnostic tool for healthcare use.

## Acknowledgements

I would like to thank my supervisor Eija Nissilä for all the time and effort put in to helping me and supporting me through completing the study, and my secondary supervisor Päivi Saavalainen for helping me with the subject and allowing me this exceptional opportunity to carry out the work in her lab. I would also like to thank my other colleagues in Päivi Saavalainen lab for helping me through general issues concerning the analysis part in this study. Additionally, I would like to thank Anne Salonen, Tinja Kanerva and Roosa Jokela in Willem M. de Vos research group in the Faculty of Medicine (Human Microbiome Research) for collaboration and expertise in microbiome research. Finally, I would like to thank Alli Hietikko and Tommy Holmberg for reading and commenting the thesis.

## References

- [1] **Baumgart DC, Carding SR.** Inflammatory bowel disease: cause and immunobiology. *Lancet* 2007;369:1627–1640.
- [2] **Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W *et al.*** Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* 2017;390:2769–2778.
- [3] **Chi KR.** Epidemiology: rising in the east. *Nature* 2016;540:100–102.
- [4] **Duchmann R, Neurath MF, zum Büschenfelde KM.** Responses to self and non-self intestinal microflora in health and inflammatory bowel disease. *Res Immunol* 1997;148:589–594.
- [5] **Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP *et al.*** Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;491:119–124.
- [6] **Ardizzone S, Cassinotti A, Manes G, Porro GB.** Immunomodulators for all patients with inflammatory bowel disease? *Therap Adv Gastroenterol* 2010;3:31–42.
- [7] **Rutgeerts P, Sandborn WJ, Feagan BG, Reinisch W, Olson A *et al.*** Infliximab for induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 2005;353:2462–2476.
- [8] **Murch SH, Lamkin VA, Savage MO, Walker-Smith JA, MacDonald TT.** Serum concentrations of tumour necrosis factor in childhood chronic inflammatory bowel disease. *Gut* 1991;32:913–917.
- [9] **Knight DM, Trinh H, Le J, Siegel S, Shealy D *et al.*** Construction and initial characterization of a mouse-human chimeric anti-TNF antibody. *Mol Immunol* 1993;30:1443–1453.

- [10] **Nielsen OH, Ainsworth MA.** Tumor necrosis factor inhibitors for inflammatory bowel disease. *N Engl J Med* 2013;369:754–762.
- [11] **Whitman WB, Coleman DC, Wiebe WJ.** Prokaryotes: the unseen majority. *Proc Natl Acad Sci* 1998;95:6578–6583.
- [12] **Sender R, Fuchs S, Milo R.** Revised estimates for the number of human and bacteria cells in the body. *PLOS Biol* 2016;14:1–14.
- [13] **Richard ML, Sokol H.** The gut mycobiota: insights into analysis, environmental interactions and role in gastrointestinal diseases. *Nat Rev Gastroenterol Hepatol* 2019.
- [14] **Main J, McKenzie H, Yeaman GR, Kerr MA, Robson D et al.** Antibody to *Saccharomyces cerevisiae* (baker's yeast) in Crohn's disease. *BMJ* 1988;297:1105.
- [15] **D'Haens G, Ferrante M, Vermeire S, Baert F, Noman M et al.** Fecal calprotectin is a surrogate marker for endoscopic lesions in inflammatory bowel disease. *Inflamm Bowel Dis* 2012;18:2218–2224.
- [16] **Fukunaga S, Kuwaki K, Mitsuyama K, Takedatsu H, Yoshioka S et al.** Detection of calprotectin in inflammatory bowel disease: Fecal and serum levels and immunohistochemical localization. *Int J Mol Med* 2017;41:107–118.
- [17] **Huseyin CE, Rubio RC, O'Sullivan O, Cotter PD, Scanlan PD.** The fungal frontier: a comparative analysis of methods used in the study of the human gut mycobiome. *Fron Microbiol* 2017;8:1–15 .
- [18] **Salonen A, Nikkilä J, Jalanka-Tuovinen J, Immonen O, Rajilić-Stojanović M et al.** Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *J Microbiol Methods* 2010;81:127–34.
- [19] **Yu Z, Morrison M.** Improved extraction of PCR-quality community DNA from digesta and fecal samples. *BioTechniques* 2004;36:808–812.
- [20] **Yu Z, Yu M, Morrison M.** Improved serial analysis of V1 ribosomal sequence tags (SARST-V1) provides a rapid, comprehensive, sequence-based characterization of bacterial diversity and community composition. *Environ Microbiol* 2006;8:603–611.
- [21] **Yu Z, Garcia-Gonzalez R, Schanbacher FL, Morrison M.** Evaluations of different hypervariable regions of archaeal 16S rRNA genes in profiling of methanogens by archaea-specific PCR and denaturing gradient gel electrophoresis. *Appl Environ Microbiol* 2008;74:889–893.
- [22] **White TJ, Bruns T, Lee S, Taylor JW.** Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: A Guide to Methods and Applications* 1990;18:315–322.
- [23] **Illumina.** 16S metagenomic sequencing library preparation guide. 2014.
- [24] **Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD.** Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 2013;79:5112–5120.

- [25] **Korpela K.** Mare: microbiota analysis in R easily. R package version 1.0. 2016. <https://github.com/katrikorpela/mare>
- [26] **R Core Team.** R: a language and environment for statistical computing [Internet]. Vienna, Austria; 2016. Available: <https://www.r-project.org/>
- [27] **Edgar RC.** Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;26:2460–2461.
- [28] **UNITE Community (2019).** UNITE general FASTA release for Fungi. Version 18.11.2018. UNITE Community.
- [29] **Simpson EH.** Measurement of diversity. *Nature* 1949;163: 688–688.
- [30] **Bazin T, Hooks KB, Barnette T, Truchetet ME, Enaud R et al.** Microbiota composition may predict anti-tnf alpha response in spondyloarthritis patients: an exploratory study. *Sci Rep* 2018;8:5446.
- [31] **Zhou Y, Xu ZZ, He Y, Yang Y, Liu L et al.** Gut microbiota offers universal biomarkers across ethnicity in inflammatory bowel disease diagnosis and infliximab response prediction. *MSystems* 2018;3:e00188-17.
- [32] **Sokol H, Leducq V, Aschard H, Pham H-P, Jegou S et al.** Fungal microbiota dysbiosis in IBD. *Gut* 2017;66:1039–1048.
- [33] **Ott SJ, Kühbacher T, Musfeldt M, Rosenstiel P, Hellmig S et al.** Fungi and inflammatory bowel diseases: Alterations of composition and diversity. *Scand J Gastroenterol* 2008;43:831–841.
- [34] **Hoffmann C, Dollive S, Grunberg S, Chen J, Li H et al.** Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. *PLoS One* 2013;8:e66019.
- [35] **Huseyin CE, O'Toole PW, Cotter PD, Scanlan PD.** Forgotten fungi—the gut mycobiome in human health and disease. *FEMS Microbiol Rev* 2017;41:479–511.
- [36] **Richard ML, Lamas B, Liguori G, Hoffmann TW, Sokol H.** Gut fungal microbiota: the yin and yang of inflammatory bowel disease. *Inflamm Bowel Dis* 2015;21:656–665.
- [37] **Baert FJ, D'Haens GR, Peeters M, Hiele MI, Schaible TF et al.** Tumor necrosis factor alpha antibody (infliximab) therapy profoundly down-regulates the inflammation in Crohn's ileocolitis. *Gastroenterology* 1999;116:22–28.
- [38] **Haga HJ, Brun JG, Berntzen HB, Cervera R, Khamashta M et al.** Calprotectin in patients with systemic lupus erythematosus: Relation to clinical and laboratory parameters of disease activity. *Lupus* 1993;2:47–50.

- [39] **Soyfoo MS, Roth J, Vogl T, Pochet R, Decaux G.** Phagocyte-specific S100A8/A9 protein levels during disease exacerbations and infections in systemic lupus erythematosus. *J Rheumatol* 2009;36:2190–2194.
- [40] **Pepper RJ, Hamour S, Chavele KM, Todd SK, Rasmussen N *et al.*** Leukocyte and serum S100A8/S100A9 expression reflects disease activity in ANCA-associated vasculitis and glomerulonephritis. *Kidney Int* 2013;83:1150–1158.
- [41] **Allali I, Arnold JW, Roach J, Cadenas MB, Butz N *et al.*** A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. *BMC Microbiology* 2017;17.
- [42] **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD *et al.*** QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7:335–336.
- [43] **Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ *et al.*** DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;13:581–583.
- [44] **Dollive S, Peterfreund GL, Sherrill-Mix S, Bittinger K, Sinha R *et al.*** A tool kit for quantifying eukaryotic rRNA gene sequences from human microbiome samples. *Genome Biol* 2012;13:R60.
- [45] **White JR, Maddox C, White O, Angiuoli SV, Fricke WF.** CloVR-ITS: automated internal transcribed spacer amplicon sequence analysis pipeline for the characterization of fungal microbiota. *Microbiome* 2013;1.
- [46] **Virtanen S, Kalliala I, Nieminen P, Salonen A.** Comparative analysis of vaginal microbiota sampling using 16S rRNA gene analysis. *PloS One* 2017;12:e0181477.
- [47] **Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z *et al.*** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402.
- [48] **Clemente JC, Ursell LK, Parfrey LW, Knight R.** The impact of the gut microbiota on human health: an integrative view. *Cell* 2012;148:1258–1270.